# OceanSITES Data Providers' Guide

Draft: Version 1.1
January 2, 2015

**Document History:**

| Date | Version | Author | Description |
|---|---|---|---|
| 2012-08-09 | | Jing Zhou | Initial Draft |
| 2013-05-30 | | Matthias Lankhorst | Edits during DMT meeting, Seoul |
| 2013-12-04 | | Nan Galbraith | renamed |
| 2013-12-27 | | Nan Galbraith | added some steps and changed reference to Data Format Specification (formerly users manual, data manual) |
| 2015-01-02 | 1.1 | Julie Thomas | Added mission statement |

# 1. OceanSITES Overview

**Mission**: The mission of OceanSITES is to collect, deliver and promote the use of high-quality data from long-term, high-frequency observations at fixed locations in the open ocean. OceanSITES typically aim to collect multidisciplinary data worldwide from the full-depth water column as well as the overlying atmosphere.

**Rationale**: Time series observations at critical or representative locations are one essential element of a global ocean observing system to complement a range of other approaches. They can provide: a unique view of the full temporal behavior of a system; accurate reference and long-time baseline data; and the maximum possible range of interlinked variables from the seafloor to the atmosphere while enabling shared resources.

All OceanSITES data are publicly available. More information about the project is available at: http://www.oceansites.org

## 1.1 Technical Documentation available

Technical documentation of the OceanSITES system consists of three pieces

**OceanSITES Data Format Reference Manual**
For data producers and users, formerly called the OceanSITES Users Manual,
this document contains a description of the OceanSITES NetCDF specification, code lists for required variables, and OceanSITES Data Management terms of reference.

**OceanSITES Data Users' Guide**
For data users, formerly called How to Access OceanSITES Data, this document
contains an outline of GDAC data directory structure and ftp/opendap access, data use policy/license, list of sites, DACs, etc. It should be used in conjunction with the Data Format Reference Manual.

**OceanSITES Data Providers' Guide**
For data producers: DACs and PIs, combining the documents How to Work with GDAC and, possibly, How to Become an OceanSITE (although that's directed more at PIs?)
This document contains guidelines for providing metadata and data, file naming scheme, how to upload to GDACs. It should be used in conjunction with the Data Format Reference Manual.

# 2. Data Management Overview

The OceanSITES data flow is carried out through three organizational units: Principal Investigators (PI), Data Assembly Centers (DAC), and Global Data Assembly Centers (GDAC).

Generally, OceanSITES publishes the "best" version available of each dataset, and does not maintain

older versions once a given version has been superseded. Data is archived at the US National Ocean Data Center, NODC.

The OceanSITES data management team has developed an implementation of NetCDF based on the community-supported Climate and Forecast (CF) standard, which supplies a standard vocabulary and some metadata conventions. The OceanSITES specification includes some requirements beyond the CF standard, for better data discovery and provenance information. In general, the PI provides the data and metadata information to a DAC; a PI may act as his/her own DAC. The DAC checks the metadata, and converts data and metadata into an OceanSITES file, and passes it on to a GDAC. The GDAC puts the files on its publicly accessible ftp server, and within 24 hours adds the file to its data catalog, and sends the file to the second GDAC to be added to its server.

# 3. About this document

This document provides guidelines for preparing data for inclusion in the OceanSITES data repository; the OceanSITES Data Format Reference Manual (or Users' Guide) contains the actual formatting and content requirements for data files. Further, this document describes QA/QC procedures to ensure the highest quality data, and describes how to work with a GDAC to distribute data. It provides detailed information on what is expected from a data provider.

# 4. Steps in adding a site to OceanSITES

The following steps describe the basic approach for a PI or DAC to contribute data to the OceanSITES Project.

## 4.1.1 Register with OceanSITES, Obtain Site and Platform Codes

A prospective PI may approach the OceanSITES project to become a member. Contact information for the project is available on the project website, OceanSITES.org. Projects with sustained, high quality time series observations in open ocean locations are the main focus of the OceanSITES project.

Once approved, the PI works with the OceanSITES project office to determine meaningful site codes and platform codes for his/her platforms. An OceanSITES site is a defined geographic location where sustained oceanographic, meteorological or other observations are made. An OceanSITES platform is an independently deployable package of instruments and sensors forming part of the site. There can be several platforms (e.g. several moorings) at a single site. The OceanSITES catalogue maintained by the project office has a full list of site codes and platform codes.

### 4.1.2  Collect Data and Metadata

The OceanSITES Data Format Reference Manual documents a fairly rigorous content standard; we hope to have fully self-documenting data in the project, to ensure the maximum usefulness of the data in the long term.

Please include all metadata required by the specification, but also record any relevant additional information to the best of your ability. Items such as instrument manufacturer and model name, sensor specifications, processing steps (e.g. source of and value of magnetic declination corrections), and conventions (e.g. whether wind is in meteorological or oceanographic convention) should be included. Note any shortcomings in the collection methods or processing.

### 4.1.3  Generate Data Files

The DAC (or PI) assembles the data into files that comply with the OceanSITES Data Format Reference Manual, available on the OceanSITES web site. The OceanSITES Data Management Team can be contacted with formatting or content standard questions. A number of software tools can be provided upon request to assist with the data file assembly, such as example files and scripts to generate example files.

The GDACs use metadata within the NetCDF files to generate an inventory 'index' file of our data holdings; the following details of the specification are used in that process, and so are of particular importance:

- **File Names**

    A file naming convention assists us in organizing our data so it can be catalogued and discovered more easily. An OceanSITES NetCDF file name is in the form;
    OS_PPP_DDD_M<_ PI>.nc

    OS – OceanSITES prefix
    PPP – Platform code from the OceanSITES catalogue
    DDD – Deployment code (PI-defined unique code for deployment)
    M – Data Mode (R: remote or near real-time collected data, P: provision data, D: delayed mode data via recovered instruments, M: mixed delayed mode and remote collected data)
    <PI> – An optional PI defined field for identification of data
    .nc – Required suffix for all files

    **Note:** no spaces or underscores are allowed in the PI defined field.

- **Global attributes**

    The GDACs maintain a data catalog that depends upon specific global attributes in the NetCDF files. The following global attributes are required for that catalog:

*site_code*: site code from the OceanSITES catalogue
*platform_code* : platform code from the OceanSITES catalogue
*date_update*: file update or creation date
*time_coverage_start*: start date of the data
*time_coverage_end*: final date of the data in
*geospatial_lat_min*: southernmost latitude
*geospatial_lat_max*: northernmost latitude
*geospatial_lon_min*: westernmost longitude between
*geospatial_lon_max*: easternmost longitude between
*geospatial_vertical_min*: minimum depth in meters for measurements
*geospatial_vertical_max*: maximum depth in meters for measurements
*update_interval*: update interval for the file (void, 'P1D')
*data_mode*: data mode (the same as appeared in the file name)

**Notes:**
- Dates and times are in UTC. The string representation of time must be "YYYY-MM-DDThh:mm:ssZ" per ISO 8601.
- update_interval is in ISO 8601: 'void' for no scheduled updates, 'P1D' for daily, or 'PT12H' for twice daily updates.
- Latitudes are specified as +/- 90 degrees; longitudes are +/- 180.

## 4.1.4 Arrange for Data Upload, Upload Data

Before submitting data to a GDAC for inclusion in the OceanSITES repository, the data provider must request an account from one of the GDACS. The GDACs are at IFREMER, in France, and at NDBC in the USA; they run scheduled data synchronization, so data only needs to be uploaded to one site.

Typically, users upload data via FTP to a GDAC-provided directory. For near real-time data, files may be updated on a daily basis. Files are not placed directly into the public FTP directory, but into users' areas; the GDACs populate the public FTP structure on a scheduled basis.

The top public FTP directory at NDBC is located at
ftp://data.ndbc.noaa.gov/data/oceansites/
The top public FTP directory at Ifremer is located at
ftp://ftp.ifremer.fr/ifremer/oceansites/

## 4.1.5 Check Status

Within 24 hours of data upload, new files will appear in the first GDAC's public FTP, if processing is successful. The second GDAC server will synchronize the data files within the following 24 hours. PIs should check that their files appear on the servers; they may also verify the contents, if desired, by downloading the released data files from the FTP servers.

The OceanSITES GDACs provide an inventory of data holdings in an 'index file' named oceansites_index.txt, in the top directory of the FTP site. Within 24 hours of data upload, there will be a line entry in the index file for each successfully processed data file. The index file is a comma-delimited text file; it is the responsibility of the data provider to check his data files in this index, to look for metadata errors.

The first field is the path from the top FTP directory to the corresponding data file. A typical OceanSITES data file would have a path like:
*DATA/SiteCode/FileName.nc*
Successive fields provide basic metadata about the file contents, taken from the metadata fields listed earlier. Please check the geo-spatial-temporal extents and the CF standard names of the data variables for errors, and discuss any problems with the GDAC or the Data Management Team.

## 4.2    Data File Changes and removals

If a file needs to be updated, the DAC can upload the new version via FTP, and the GDAC processing will replace the existing file on the public server with the new version within 24 hours. You may check the index file to be sure the update was successful.

DACs do not have the ability to directly remove data files from the public-facing FTP directories. To remove a data file from the GDAC FTP severs, simply upload a zero-length data file with the intended file name to the respective FTP account. The GDAC processing programs will remove the corresponding data file from the GDAC FTP server within 24 hours, and from the other GDAC FTP server within the next 24 hours after synchronization.

Note that real time data should be replaced with delayed-mode data as soon as the latter is available. To achieve this, simply upload an empty file with the same name as your real time data file, when you're uploading your delayed-mode data.

# 5    Metadata and Quality Issues

## 5.1    Documenting instruments used

## 5.2    Documenting calibrations applied

## 5.3    Documenting processing done